03-02-00

A

jc715 U.S. PTO
03/01/00

Certificate of Mailing

Date of Deposit March 1, 2000 Label Number: EL488650596US

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as **"Express Mail Post Office to Addressee"** with sufficient postage on the date indicated above and is addressed to: BOX PATENT APPLICATION, Assistant Commissioner for Patents, Washington, D.C. 20231.

Luis Cruz
Printed name of person mailing correspondence

Signature of person mailing correspondence

jc678 U.S. PTO
09/516061
03/01/00

UTILITY PATENT APPLICATION TRANSMITTAL UNDER 37 CFR §1.53(b)

Attorney Docket Number	50093/016001
Applicant	VENKAT GOPALAN, MILAN JOVANOVIC, PAUL S. EDER, TONY GIORDANO, GORDON D. POWERS and K. ASISH XAVIER
Title	NOVEL BACTERIAL RNASE P PROTEINS AND THEIR USE IN IDENTIFYING ANTIBACTERIAL COMPOUNDS

PRIORITY INFORMATION:

APPLICATION ELEMENTS:

Cover sheet	1 pages
Specification	26 pages
Claims	2 pages
Abstract	1 pages
Drawing	8 pages
Combined Declaration and POA, which is: ☒ Unsigned; ☐ Newly signed for this application; ☐ A copy from prior application [**SERIAL NUMBER**] and the entire disclosure of the prior application is considered as being part of the disclosure of this new application and is hereby incorporated by reference therein.	3 pages
Statement Deleting Inventors	0 pages
Sequence Statement	0 pages
Sequence Listing on Paper	0 pages
Sequence Listing on Diskette	0 pages
Small Entity Statement, which is: ☐ Unsigned; ☐ Newly signed for this application; ☐ A copy from prior application [**SERIAL NUMBER**] and such small entity status is still proper and desired.	0 pages
Preliminary Amendment	0 pages

IDS	0 pages
Form PTO 1449	0 pages
Cited References	0 pages
Recordation Form Cover Sheet and Assignment	0 pages
Assignee's Statement	0 pages
English Translation	0 pages
Certified Copy of Priority Document	0 pages
Return Receipt Postcard	1
FILING FEES:	
Basic Filing Fee: $345	$345.00
Excess Claims Fee:14 - 20 x $9	$0
Excess Independent Claims Fee:3 - 3 x $39	$0
Multiple Dependent Claims Fee: $260/$130	$0
Total Fees:	$345.00

- ☒ Enclosed is a check for $345.00 to cover the total fees.
- ☐ Charge [**AMOUNT**] to Deposit Account No. 03-2095 to cover the total fees.
- ☐ The filing fee is not being paid at this time.
- ☒ Please apply any other charges, or any credits, to Deposit Account No. 03-2095.

CORRESPONDENCE ADDRESS:

Kristina Bieker-Brady, Ph.D.
Reg. No. 39,109
Clark & Elbing LLP
176 Federal Street
Boston, MA 02110

Telephone: 617-428-0200
Facsimile: 617-428-7045

K. Bieker-Brady
Signature

March 1, 2000
Date

\\Ceserver\documents\50093\50093.016001 Appln Trnsmtl Ltr.wpd

Certificate of Mailing

Date of Deposit March 1, 2000

Label Number: EL488650596US

I hereby certify under 37 CFR 1.10 that this correspondence is being deposited with the United States Postal Service as **"Express Mail Post Office to Addressee"** with sufficient postage on the date indicated above and is addressed to BOX PATENT APPLICATION, Assistant Commissioner of Patents, Washington, D.C. 20231.

Luis Cruz
Printed name of person mailing correspondence

Signature of person mailing correspondence

APPLICATION

FOR

UNITED STATES LETTERS PATENT

APPLICANT	:	VENKAT GOPALAN, MILAN JOVANOVIC, PAUL S. EDER, TONY GIORDANO, GORDON D. POWERS and K. ASISH XAVIER
TITLE	:	NOVEL BACTERIAL RNASE P PROTEINS AND THEIR USE IN IDENTIFYING ANTIBACTERIAL COMPOUNDS

\\Ceserver\documents\50093\50093.016001 Appln Cover Sheet.wpd

NOVEL BACTERIAL RNase P PROTEINS AND THEIR USE IN IDENTIFYING ANTIBACTERIAL COMPOUNDS

Background of the Invention

This invention relates to novel bacterial ribonuclease P protein subunits and their use as targets in screening assays to identify compounds useful as antibacterial agents.

Ribonuclease P (RNase P) is an endoribonuclease that cleaves the 5'-terminal leader sequences of precursor tRNAs. RNase P has been characterized in representative a number of species.

In bacteria, the structure of the RNase P holoenzyme is composed of a catalytic RNA subunit (350-450 nucleotides; encoded by the *rnp* B gene) and a single protein subunit (110-160 amino acids; encoded by the *rnp* A gene); both are essential for *in vivo* activity. In *Escherichia coli* (*E. coli*) the RNA subunit is termed M1 and the protein subunit is C5. The C5 protein engages in specific interactions with the M1 RNA to stabilize certain M1 RNA conformations. Through these interactions with M1, C5 plays a critical role in the recognition/binding of some substrates.

Comparison of RNase P protein subunits between bacterial species reveals that their primary structures have only a moderate degree of identity. For example, the protein subunits of *Bacillus subtilis* (*B. subtilis*) and *E. coli* are 30% identical. The functional significance of some conserved amino acid residues has been confirmed by mutagenesis studies which have shown that these conserved amino acids play a significant role in the catalytic function of the RNase P holoenzyme.

The tertiary structure of the RNase P protein subunit expressed in *B. subtilis* has been determined by X-ray crystallography. The overall topology of α -helices and β -sheets is $\alpha1$ $\beta1$ $\beta2$ $\beta3$ $\alpha2$ $\beta4$ $\alpha3$, with an uncommon $\beta3$ $\alpha2$ $\beta4$ cross-over connection that may confer specific functional consequences. Another functional aspect of the protein is the long loop connecting $\beta2$ to $\beta3$, termed the metal binding loop, which binds Zn^{2+} ions and mediates interlattice contacts. In addition, the crystal structure reveals an overall fold that is similar to the ribosomal protein S5, translational elongation factor EF-G (domain IV), and DNA gyrase.

Many pathogens exist for which there are few effective treatments and the number of strains resistant to available drugs is continually increasing. Accordingly, novel compositions and methods for assaying RNase P function would be useful for identifying antimicrobial compounds against these pathogens.

Summary of the Invention

Certain RNase P amino acid positions are markedly conserved, as revealed by comparing the protein subunit sequences using the ClustalW multiple alignment program indicating that the residues may be important in RNase P function. The invention features novel polypeptides related to the protein component of the RNase P holoenzyme in several pathogenic bacterial species, as well as the nucleic acid sequences which encode these proteins. The invention also features methods of using these sequences identify additional RNase P nucleic acids and proteins, and methods to screen for compounds which inhibit the RNase P function. Such compounds can be used as antibacterial agents.

In the first aspect, the invention features an isolated polypeptide comprising an RNase P consensus sequence wherein said polypeptide has RNase P protein activity. In a preferred embodiment of this aspect, the polypeptide comprises an amino acid sequence selected from SEQ ID NOS: 20-38.

In the second aspect, the invention features an isolated nucleic acid sequence, wherein the sequence encodes a polypeptide comprising an amino acid

sequence substantially identical to an amino acid sequence containing an RNase P consensus and has RNase P protein activity. In preferred embodiments, the sequence encodes a polypeptide comprising an amino acid sequence selected from SEQ ID NOS: 20-38 and/or the sequence is selected from SEQ ID NOS: 1-19.

In the third aspect, the invention features a transgenic host cell including a heterologous nucleic acid sequence encoding the polypeptide of the first aspect of the invention.

In the fourth aspect, the invention features an antibody that specifically binds to the polypeptide having SEQ ID NOS:20-38 of the first aspect of the invention.

In the fifth aspect, the invention features a method of identifying an antibiotic agent, said method including: i) obtaining an RNase P holoenzyme comprising the polypeptide of the first aspect of the invention; ii) contacting the holoenzyme with an RNase P substrate in the presence and in the absence of a compound; and iii) measuring the enzymatic activity of the holoenzyme; wherein a compound is identified as an antibiotic agent if said compound produces a detectable decrease in said RNase P enzymatic activity as compared to activity in the absence of the compound. In various preferred embodiments, the polypeptide is substantially identical to a polypeptide of SEQ ID NOS:20-38, the activity is measured by fluorescence spectroscopy, the RNase substrate is fluorescently tagged $ptRNA^{Gln}$, the fluorescence analysis is carried out in a buffer comprising 10-40 mg/ml carbonic anhydrase and 10-100 μg/ml polyC, or the buffer further includes at least one of the following: 0.5-5% glycerol; 10-100 μg/ml hen egg lysozyme; 10-50 μg/ml tRNA; or 1-10 mM DTT.

In the sixth aspect, the invention features a method for making a $ptRNA^{Gln}$ that includes annealing two RNA fragments together by heating to about 65°C to about 80°C for about 5 minutes, followed by cooling to 20-25° C.

The term "nucleic acid" encompasses both RNA and DNA, including cDNA, genomic DNA, complementary antisense nucleic acids capable of

decreasing RNase P activity, and synthetic (e.g., chemically synthesized) DNA. The nucleic acid may be double-stranded or single-stranded. Where single-stranded, the nucleic acid may be a sense strand or an antisense strand.

By "isolated nucleic acid" is meant a DNA or RNA that is separated from the coding sequences with which it is naturally contiguous (one on the 5' end and one on the 3' end) in the genome of the organism from which it is derived. Thus, in one embodiment, an isolated nucleic acid includes some or all of the 5' and/or 3' non-coding (e.g., promoter) sequences which are immediately contiguous to the coding sequence. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence.

By "isolated polypeptide" is meant a preparation which is at least 60% by weight (dry weight) the polypeptide of interest. Preferably the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight the polypeptide of interest. Purity can be measured by any appropriate standard method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis.

Moreover, an "isolated" nucleic acid or polypeptide is meant to include fragments which are not naturally occurring as fragments and would not be found in the natural state.

By "a polypeptide containing RNase P activity" is meant a polypeptide sequence that, when combined with an RNA subunit to form an RNase P holoenzyme, has 20%, 50%, 75%, or even 100% or more, of the enzymatic activity of an *E. coli* or *B. subtilis* RNase P holoenzyme. Preferably, the RNA subunit is from the same species when activity is tested. The enzymatic activity can be

assessed, for example, by measuring hydrolysis of an RNase P substrate. Standard methods for conducting such hydrolysis assays are described herein and in the literature (see, e.g., Altman and Kirsebom, Ribonuclease P, *The RNA World*, 2^{nd} Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999; Pascual and Vioque, Proc. Natl. Acad. Sci. 96: 6672, 1999; Geurrier-Takada et al., Cell 35: 849, 1983; Tallsjö and Kirsebom, Nucleic Acids Research 21: 51, 1993; Peck-Miller and Altman, J. Mol. Biol. 221: 1, 1991; Gopalan et al., J. Mol. Biol. 267: 818, 1997; and WO 99/11653).

By "RNase P substrate" is meant a substrate in which hydrolysis by an RNase P holoenzyme requires the presence of the RNase P protein subunit.

By "identity" is meant the relationship between two or more polypeptide sequences or two or more nucleic acid sequences, as determined by comparing the degree of sequence relatedness. "Identity" can be readily calculated by known methods, including but not limited to those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, and Devereux, eds., M. Stockton Press, New York, 1991; and Carillo and Lipman, SIAM J. Applied Math. 48: 1073, 1988.

Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are available in publicly available computer programs. Computer program methods to determine identity between two sequences include, but are not limited to, the GCG program package (Devereux et al., Nucleic Acids Research 12(1): 387, 1984), BLASTP, BLASTN, and FASTA (Altschul et al., J. Mol. Biol. 215: 403 (1990). The well known Smith Waterman algorithm may also be used to determine identity. The BLAST program is publicly available from NCBI and other sources

(*BLAST Manual*, Altschul, et al., NCBI NLM NIH Bethesda, MD 20894). Searches can be performed in URLs such as the following http://www.ncbi.nlm.nih.gov/BLAST/unfinishedgenome.html; or http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi.

As an illustration of percent identity, if a test nucleic acid sequence (TN) has 95% identity to a reference nucleic acid sequence (RN) at the specified bases, then TN is identical to RN at the specified bases, except that TN may include point mutations in 5% of the total number of nucleic acids present in RN. Thus, 5% of nucleic acids found in RN may be deleted or substituted with another nucleic acid. In addition, the sequence of TN may contain, as compared to the specified RN bases, insertions of nucleic acids totaling up to 5% of the nucleic acids present in RN. These mutations, as compared to the RN sequence, may occur at the 5' or 3' terminal positions or anywhere between those terminal positions, interspersed either individually among the specified nucleic acids or in one or more contiguous groups of specified nucleic acids. As in the present invention, for nucleic acids encoding proteins, trinucleotide sequences encoding the same amino acid may optionally be treated as identical.

Analogously, a test polypeptide (TP) has an amino acid sequence 95% identical to a reference amino acid sequence (RP) if TP is identical to RP at the specified amino acids, except that TP contains amino acid alterations totaling 5% of the total number of specified amino acids in RP. These alterations include deletions of amino acids or substitutions with one or more other specified amino acids. In addition, the alterations include insertions of other amino acids totaling up to 5% of the total amino acids present in the specified RP amino acids. The alterations in the TP amino acid sequence as compared to the RP sequence may occur at the amino or carboxy terminal positions, or anywhere between those terminal positions, interspersed either individually among residues or in one or more contiguous groups.

By "an RNase P consensus sequence" is meant a sequence which, when aligned to the *E. coli* RNase P sequence using the ClustalW program and performing a comparison of the specified amino acid sequences, shows conservation of at least nine of the following specified 20 amino acid residues in the *E. coli* RNase P protein subunit: R11, L12, F18, R46, G48, V51, K53, K54, A59, V60, R62, N63, K66, R67, R70, L80, D84, V86, L101, and L105. Preferably, the consensus sequence conserves at least 13 of the 20 residues. It is also preferred that the aligned consensus sequence contain at least seven of the following subset of nine amino acid residues in the *E. coli* RNase P protein: F18, R46, K53, A59, R62, N63, K66, R67, R70, more preferably, at least eight of the amino acids, and, most preferably, all nine amino acids of the above subset. For the purpose of determining identity in the present invention, identity of amino acids or other than those for which the amino acid is specified in the consensus sequence are ignored in the comparison when calculating identity of nucleic acids encoding an RNase P consensus sequence degenerate codons encoding the designated amino acid are treated as identical.

The RNase P sequences claimed as part of the present invention specifically exclude those sequences in the RNase P database (James W. Brown, The Ribonuclease P Database, Nucleic Acids Research 27(1):314 (1999)) as posted on the internet on March 1, 2000. Also excluded are the RNase P polypeptide and nucleic acids described by nucleic acid or amino acid sequence in EP 0811 688 A2 (*Staphylococcus aureus*) and WO 99/11653 (*S. pneumoniae*).

A "substantially identical" RNase P sequence is one which has or encodes a polypeptide having at least 95% identity, preferably 100% identity, to the twenty amino acids provided from the sequence of *E. coli* RNase P hereinbefore above.

"Transformation" or "transfection" means any method for introducing foreign molecules, such as nucleic acids, into a cell. Lipofection, DEAE-dextran-mediated transfection, microinjection, protoplast fusion, calcium phosphate precipitation, retroviral delivery, electroporation, and biolistic transformation are

just a few of the methods known to those skilled in the art which may be used. These techniques may be applied to the transformation or transfection of a wide variety of cell types and intact tissues including, without limitation, intracellular organelles (e.g., and mitochondria and chloroplasts), bacteria, yeast, fungi, algae, animal tissue, and cultured cells.

By "transgenic host cell" means a cell (or a descendent of a cell) transformed or transfected with a heterologous nucleic acid sequence comprising a coding sequence operably linked to one or more sequence elements, e.g., a promoter, which directs transcription and/or translation such that the heterologous coding sequence is expressed in said host cell. The transgenic host cells may be either stably or transiently transfected.

By "operably linked" is meant that a selected nucleic acid is positioned adjacent to one or more sequence elements, e.g., a promoter, which direct transcription and/or translation of the selected nucleic acid.

By "specifically binds" is meant an antibody that recognizes and binds to the full length protein or subfragment of any one of SEQ ID NOS: 20-38, but which does not substantially recognize and bind to other molecules in a sample, including other RNase P proteins.

Other features and advantages of the invention will be apparent from the following detailed description and from the claims.

Description of the Figures

Fig. 1 shows the sequence alignment of previously known bacterial RNase P protein subunits using the ClustalW alignment program (Thompson et al., Nucleic Acids Research 22: 4673, 1994) and the alignment of the RNase P sequences of the present invention. The aligned fragments of the known RNase P sequences are designated by (*) and the aligned fragments of the RNase P sequences of the invention are designated by (#).

Figs. 2A-2S shows the nucleic acid sequences (SEQ ID NOs 1-19) encoding the amino acid sequences (SEQ ID NOs 20-38) of the bacterial RNase P polypeptides of the invention. The nucleic acid and amino acid sequences were derived from the following pathogenic bacterial species: *Streptococcus mutans* (Fig. 2A; SEQ ID NOs: 1 and 18, respectively); *Klebsiella pneumoniae* (Fig. 2B; SEQ ID NOs: 2 and 19, respectively); *Salmonella paratyphi* A (Fig. 2C; SEQ ID NOs: 3 and 20, respectively); *Pseudomonas aeruginosa* (Fig. 2D; SEQ ID NOs: 4 and 21, respectively); *Corynebacterium diphtheriae* (Fig. 2E; SEQ ID NOs: 5 and 22, respectively); *Chlamydia trachomatis* (Fig. 2F; SEQ ID NOs: 6 and 23, respectively); *Vibrio cholerae* Serotype 01, Biotype El Tor, Strain N16961 (Fig. 2G; SEQ ID NOs: 7 and 24, respectively); *Neisseria gonorrhoea* FA 1090 (Fig. 2H; SEQ ID NOs: 8 and 25, respectively); *Neisseria meningitidis* Serogroup A, Strain Z2491 (Fig. 2I; SEQ ID NOs: 9 and 26, respectively); *Streptococcus pyogenes* M1 (Fig. 2J; SEQ ID NOs: 10 and 27, respectively); *Bordetella pertussis* Tohama I (Fig. 2K; SEQ ID NOs: 11 and 28, respectively); *Porphyromonas gingivalis* W83 (Fig. 2L; SEQ ID NOs: 12 and 29, respectively); *Streptococcus pneumoniae* Type 4 (Fig. 2M; SEQ ID NOs: 13 and 30, respectively); *Clostridium difficile* 630 (Fig. 2N; SEQ ID NOs: 14 and 31, respectively); *Camphylobacter jejuni* NCTC (Fig. 2O; SEQ ID NOs: 15 and 32, respectively); *Bacillus anthracis* Ames (Fig. 2P; SEQ ID NOs: 16 and 33, respectively); *Mycobacterium avium* 104 (Fig. 2Q; SEQ ID NOs: 17 and 34, respectively); *Staphylococcus aureus* NCTC 8325 (Fig. 2R; SEQ ID NOs: 18 and 35, respectively); and *Staplylococcus aureus* COL (Fig. 2S; SEQ ID NOs: 19 and 36, respectively).

Detailed Description

The invention features novel polypeptides that form the protein component of the RNase P holoenzyme in several pathogenic bacterial species, as well as the nucleic acid sequences which encode these proteins. The invention also features methods of using these sequences to form the protein subunit of RNase P

holoenzymes to screen for compounds which inhibit the function of the holoenzymes. Such inhibitory compounds can be used as anti-bacterial agents.

1. Identification of the Novel RNase P Protein Subunits

The novel RNase P amino acid and nucleic acid sequences were discovered using the following strategy. First, the genomic databases of several pathogenic bacteria were searched using the BLAST program (Altschul et al., J. Mol. Bio. 215: 403, 1990) and known RNase P polypeptide sequences from *E. coli* (gram-negative) and *B. subtilis* (gram-positive) as "query" sequences. Given that the largest number of known RNase P protein subunit sequences correspond to sequences from gram-negative and gram-positive bacteria, "query" sequences from both bacterial groups were used in the search to ensure that all novel sequences having homology to known RNase P sequences would be identified.

BLAST searches of genomic databases for potential RNase P homologues were performed in the following URLs: http://www.ncbi.nlm.nih.gov/BLAST/ unfinishedgenome. html; and http://www.tigr.org/cgi-bin/BlastSearch/blast.cgi.

The BLAST program only considered hits with a P-value of less than or equal to 10^{-5} to ensure that random hits were not sampled.

The above-described searches often yielded multiple hits in the genomic databases. To identify which sequences were genuine RNase P protein subunits, we determined whether the sequences also contained an RNase P consensus sequence, which we defined as a sequence that, upon alignment with known RNase P sequences using the ClustalW program, conserves at least nine of the following twenty amino acids in the *E. coli* RNase P protein sequence: R11, L12, F18, R46, G48, V51, K53, K54, A59, V60, R62, N63, K66, R67, R70, L80, D84, V86, L101, and L105. Preferred sequences contained at least thirteen out of the twenty residues and/or had at least seven of the following amino acid subset: F18, R46, K53, A59, R62, N63, K66, R67, and R70.

This RNase P consensus sequence was derived as follows. We aligned the sequences of the known bacterial RNase P protein subunits using the ClustalW

alignment program (Thompson et al., *supra*) (see Fig. 1, the previously known RNase P sequences are designated by (*) and were obtained from the RNase P database; www.jwbrown.mbio.ncsu.edu/rnasp/home.html.) This ClustalW alignment was then manually refined to align highly conserved RNase P hydrophobic and basic residues that had been demonstrated by mutation studies to be important for RNase P catalytic function (Gopalan et al., J. Mol. Biol. 267: 818, 1997). The spacing between the conserved residues, as well as the identity of the individual residues, appears critical to RNase P function.

Twenty amino acids were identified as highly conserved (shown as the shaded residues in Fig. 1). The percent of RNase P sequences which conserve each of the shaded residues is shown below the sequence information as percent identity. Based upon these known sequences, we determined that a polypeptide identified by our above-described RNase P BLAST search contained an RNase consensus sequence and was a genuine RNase P protein subunit if it contained at least nine of the above-described twenty amino acids. Preferred polypeptides have a consensus sequence with at least 13 of the amino acids and/or conserve at least seven of the following subset of amino acids: F18, R46, K53, A59, R62, N63, K66, R67, and R70. This subset of amino acids is preferred because it has been identified as playing a significant role in RNase P function through mutation studies (Gopalan et al., J. Mol. Biol. 267: 818 1997) and the determination of the RNase P three dimensional structure (Stams et al., Science 280: 752, 1998). As shown in Fig. 2, the three dimensional structure reveals that all of the residues that make up the above-described nine amino acid subset are proximal to each other in the tertiary structure of the protein, despite the distance between some of the residues in the primary structure.

2. RNase P Protein Amino Acid and Nucleic Acid Sequences

The novel RNase P proteins of the invention, and the nucleic acid sequences which encode the proteins, are derived from the following bacterial species: *Streptococcus mutans* UAB159; *Klebsiella pneumoniae* M6H 78578;

Salmonella paratyphi A (ATCC 9150); *Pseudomonas aeruginosa* PAO1; *Corynebacterium diphtheriae*; *Chlamydia trachomatis* MoPn; *Vibrio cholerae* Serotype 01, Biotype El Tor, Strain N16961; *Neisseria gonorrhoea* FA 1090; *Neisseria meningitidis* Serogroup A, Strain Z2491; *Streptococcus pyogenes* M1; *Bordetella pertussis* Tohama I; *Porphyromonas gingivalis* W83; *Streptococcus pneumoniae* Type 4; *Clostridium difficile* 630; *Camphylobacter jejuni* NCTC; *Bacillus anthracis* Ames; *Mycobacterium avium* 104. *Staphylococcus aureus* NCTC 8325; and *Staplylococcus aureus* COL. The sequences are shown in Fig. 2.

All of the novel RNase P protein sequences were identified by the above-described BLAST search. The alignment of these sequences with the known RNase P sequences is also shown in Fig. 1 (the RNase P sequences of the present invention are designated by (#)). This alignment demonstrates that the amino acid sequences of the invention all contain RNase P consensus sequences. Therefore, these polypeptides are genuine RNase P proteins.

The RNase P identification is further supported by the protein structure of the polypeptides of the invention, as determined by SWISS-MODEL. The SWISS MODEL is an automated protein modelling server running at the Glaxo Wellcome Experimental Research in Geneva, Switzerland (http://www.expasy.ch/swissmod/swiss.model). The polypeptide sequences of the invention were readily folded (at least in part) into the tertiary structure of the *B. subtilis* RNase P protein subunit (Stams et al., *supra*). It is noteworthy that conserved residues in the newly identified sequences are modeled into positions which are spatially and structurally identical to the RNase P protein subunit of *B. subtilis*.

Further support for the RNase P identification is as follows. Using the above-described BLAST search and consensus sequence determination, we independently identified the sequence for an RNase P protein subunit from the genomic database of *Staphylococcus aureus* (*S. aureus*). Although this sequence had been previously identified as an RNase P protein subunit and its RNase P

activity had been confirmed by assay (EPA 0 811 688 A2), our independent discovery of this RNase P sequence provides proof of principle that our method of searching for RNase P protein subunits predictably identifies polypeptides that have RNase P activity.

The invention features purified or isolated RNase P protein subunits. As used herein, both "protein" and "polypeptide" mean any chain of amino acids, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation). Thus, the term RNase P protein subunit includes full-length, naturally-occurring RNase P proteins, preproteins, and proproteins, as well as recombinantly or synthetically produced polypeptides that correspond to full-length, naturally-occurring RNase P proteins or to particular domains or portions of naturally-occurring proteins. These proteins are produced using standard techniques (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1995; Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985 (1987 Suppl.); and Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Preferred RNase P proteins include a sequence substantially identical to all or a portion of a naturally occurring RNase P protein subunit, e.g., including all or a portion of any of the sequences shown in Fig. 2 (SEQ ID NOS: 20-38).

In the case of polypeptide sequences which are less than 100% identical to a reference sequence, the non-identical positions are preferably, but not necessarily, conservative substitutions for the reference sequence. Conservative substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine, and leucine; aspartic acid and glutamic acid; asparagine and glutamine; serine and threonine; lysine and arginine; and phenylalanine and tyrosine.

Preferred polypeptides are those which are soluble under normal physiological conditions. Also within the invention are soluble fusion proteins in

which a full-length or subfragment of RNase P protein (e.g., one or more domains) is fused to an unrelated protein or polypeptide (i.e., a fusion partner) to create a fusion protein.

Structurally related RNase P polypeptides of the invention include, but are not limited to, polypeptides with additions or substitutions of amino acid residues within the amino acid sequence encoded by the RNase P nucleic acid sequences described herein these changes result in a silent change, thus producing a functionally equivalent gene product. Amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid.

Preferred RNase P polypeptides and variants have 20%, 50%, 75%, 90%, or even 100% or more of the activity of one of the bacterial RNase P proteins of SEQ ID NOS: 20-38 shown in Fig. 2. Such comparisons are generally based on equal concentrations of the molecules being compared. The comparison can also be based on the amount of protein or polypeptide required to reach the maximal activation obtainable.

In general, RNase P proteins according to the invention can be produced by transformation (transfection, transduction, or infection) of a host cell with all or part of a RNase P-encoding nucleic acid sequence of the present invention in a suitable expression vehicle. Such expression vehicles include: plasmids, viral particles, and phage. For insect cells, baculovirus expression vectors are suitable. The entire expression vehicle, or a part thereof, can be integrated into the host cell genome. In some circumstances, it is desirable to employ an inducible expression

vector, e.g., the LACSWITCH™ Inducible Expression System (Stratagen, LaJolla, CA).

Those skilled in the field of molecular biology will understand that any of a wide variety of expression systems can be used to provide the recombinant protein (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1995; Pouwels et al., *Cloning Vectors: A Laboratory Manual*, 1985 (1987 Suppl.); and Sambrook et al., *Molecular Cloning, A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). The precise host cell used is not critical to the invention. The RNase P protein can be produced in a prokaryotic host (e.g., *E. coli* or *B. subtilis*) or in a eukaryotic host (e.g., *Saccharomyces* or *Pichia*; mammalian cells, e.g., COS, NIH 3T3 CHO, BHK, 293, or HeLa cells; or insect cells; or plant cells).

The host cells harboring the expression vehicle can be cultured in conventional nutrient media adapted as needed for activation of a chosen gene, repression of a chosen gene, selection of transformants, or amplification of a chosen gene.

RNase P proteins can be produced as fusion proteins. For example, the expression vector pUR278 (Ruther et al., EMBO J. 2: 1791, 1983), can be used to create lacZ fusion proteins. The pGEX vectors can be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can be easily purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. The pGEX vectors are designed to include thrombin or factor Xa protease cleavage sites so that the cloned target gene product can be released from the GST moiety.

The invention also features the isolated nucleic acid sequences of SEQ ID NOS: 1-19 shown in Fig. 2, and nucleic acid sequences that encode one or more portions or domains of an RNase P protein subunit, including but not limited to the $\alpha1$, $\alpha2$, $\alpha3$, $\beta1$, $\beta2$, $\beta3$, and $\beta4$ portions of the protein.

Preferred nucleic acids encode polypeptides that are soluble under normal physiological conditions. Also within the invention are nucleic acids encoding fusion proteins in which the whole RNase P protein or a sub-fragment is fused to an unrelated protein or polypeptide (e.g., a marker polypeptide or a fusion partner) to create a fusion protein. For example, the polypeptide can be fused to a hexa-histidine tag to facilitate purification of bacterially expressed protein, or to a hemagglutinin tag to facilitate purification of protein expressed in eukaryotic cells.

The fusion partner can be, for example, a polypeptide which facilitates secretion, e.g., a secretory sequence. Such a fused protein is typically referred to as a preprotein. The secretory sequence can be cleaved by the host cell to form the mature protein. Also within the invention are nucleic acids that encode mature RNase P protein fused to a polypeptide sequence to produce an inactive proprotein. Proproteins can be converted into the active form of the protein by removal of the inactivating sequence.

The nucleic acids of the invention further include sequences that hybridize, e.g., under high stringency hybridization conditions (as defined herein), to all or a portion of the nucleic sequence of any one of SEQ ID NOS: 1-19, or any of their complements. As used herein, high stringency conditions include hybridizing at 68°C in 5x SSC/5x Denhardt solution/1.0% SDS, or in 0.5 M NaHPO4 (pH 7.2)/ 1mM EDTA/7% SDS, or in 50% formamide/0.25 M NaHPO4 (pH 7.2)/0.25 M NaCl/1 mM EDTA/7% SDS; and washing in 0.2x SSC/0.1% SDS at room temperature or at 42°C, or in 0.1x SSC/0.1% SDS at 68°C, or in 40 mM NaHPO4 (pH 7.2)/1 mM EDTA/5% SDS at 50 °C, or in 40 mM NaHPO4 (ph 7.2)/ 1 mM EDTA/1% SDS at 50°C. The parameters of salt concentration and temperature can be varied to achieve the desired level of identity between the probe and the target nucleic acid. Further guidance regarding hybridizing conditions is provided, for example, in Sambrook et al., *Molecular Cloning, A Laboratory Manual*, Cold Springs Harbor Press, NY, 1989; Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, NY, 1995).

The hybridizing portion of the hybridizing nucleic acids are preferably 20, 30, 50, or 70 bases long. Preferably, the hybridizing portion of the hybridizing nucleic acid is 80%, more preferably 95%, or even 98% identical, to the sequence of a portion or all of a nucleic acid encoding an RNase P protein subunit. Hybridizing nucleic acids of the type described above can be used as a cloning probe, a primer (e.g., a PCR primer), or a diagnostic probe. Preferred hybridizing nucleic acids encode a polypeptide having some or all of the biological activities possessed by a naturally-occurring RNase P protein subunit. Such biological activity can be determined by functional RNase P assay as described herein.

Hybridizing nucleic acids can be additional splice variants of the RNase P protein gene. Thus, they may encode a protein which is shorter or longer than the different forms of RNase P described herein. Hybridizing nucleic acids may also encode proteins that are related to RNase P (e.g., proteins encoded by genes which include a portion having a relatively high degree of identity to the RNase P genes described herein).

The invention also features vectors and plasmids that include a nucleic acid of the invention which is operably linked to a transcription and/or translation sequence to enable expression, e.g., expression vectors.

2. RNase P Antibodies

The bacterial RNase P proteins and polypeptides (or immunogenic fragments or analogs) can be used to raise antibodies useful in the invention, and such polypeptides can be produced by recombinant or peptide synthetic techniques (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1995). In general, the peptides can be coupled to a carrier protein, such as KLH, mixed with an adjuvant, and injected into a host mammal. Antibodies can be purified by peptide antigen affinity chromatography.

In particular, various host animals can be immunized by injection with an RNase P protein or polypeptide. Host animals include rabbits, mice, guinea pigs, and rats. Various adjuvants can be used to increase the immunological response,

depending on the host species, including but not limited to, Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, and potentially useful human adjuvants such as BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*.

Antibodies within the invention include polyclonal antibodies, humanized or chimeric antibodies, single chain antibodies, Fab fragments, $F(ab')_2$ fragments, molecules produced using a Fab expression library, and monoclonal antibodies.

Monoclonal antibodies, can be prepared using the RNase P proteins described above and standard hybridoma technology (see, e.g., Kohler et al., Nature 256: 495, 1975; Kohler et al., Eur. J. Immunol. 6: 511, 1976; Hammerling et al., In *Monoclonal Antibodies and T Cell Hybridomas*, Elsevier, NY, 1981; Ausubel et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1995).

In particular, monoclonal antibodies can be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture such as described in Kohler et al., Nature 256: 495, 1975, and U.S. Patent No. 4,376,110; the human B-cell hybridoma technique (Kosber et al., Immunology Today 4: 72, 1983; Cole et al., Proc. Natl. Acad. Sci. *USA* 80: 2026, 1983), and the EBV-hybridoma technique (Cole et al., *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96, 1983). Such antibodies can be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD, and any subclass thereof. The hybridoma producing the mAb of this invention can cultivated *in vitro* or *in vivo*. The ability to produce high titers of mAbs *in vivo* makes this the presently preferred method of production.

Once produced, polyclonal or monoclonal antibodies are tested for specific RNase P recognition by Western blot or immunoprecipitation analysis by standard methods, for example, as described in Ausubel et al., *Current Protocols in*

Molecular Biology, John Wiley & Sons, New York, 1995. Preferred antibodies specifically bind the RNase P proteins of the invention.

Preferably, the antibodies of the invention are produced using fragments of the RNase P protein which lie outside highly conserved regions and appear likely to be antigenic, by criteria such as high frequency of charged residues. In one specific example, such fragments are generated by standard techniques of PCR, and are then cloned into the pGEX expression vector. Fusion proteins are expressed in *E. coli* and purified using a glutathione agarose affinity matrix (Ausubel, et al., *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1995).

Another aspect the invention features a method for detecting an RNase P protein. This method includes: contacting an antibody that specifically binds an RNase P protein of the present invention to a biological sample under conditions that allow the formation of RNase P-antibody complexes; and detecting the complexes, if any, as an indication of the presence of RNase P protein in the biological sample.

3. Screening for Antibacterial Agents: Example

The rnpA genes encoding the RNase P proteins or protein subfragments of the invention are amplified from genomic DNA by established PCR methods. The amplified DNA sequences that encode the RNase P protein genes are subcloned into expression plasmids, which contain fusion sequences allowing the subcloned gene to be expressed in a transformed or transfected host cell as a "tagged" fusion protein. *E. coli* cells are transformed with the plasmid DNA, protein expression is induced, and the overexpressed fusion protein is isolated by affinity purification according to established protocols.

Each of the purified RNase P proteins is combined with a renatured cognate RNase P RNA subunit from the same, or a different, bacterial organism, under conditions that reconstitute enzymatic activity. It is possible to reconstitute a functional RNase P holoenzyme using a protein subunit and an RNA subunit from

different species (e.g., *B. subtilis, E. coli,* or *S. aureus*). The conditions for reconstitution include heat denaturing the RNA subunit then slowly cooling in a physiologically similar buffer. A buffer for folding the RNA component of RNase P is 10-50 mM Tris-HCl/MOPS/HEPES (pH=7.0-8.0), 25-500 mM KCl/NaCl/NH_4 and 1-25 mM $MgCl_2$. The RNA is heated to 65°C for 5 minuotes, 55°C for minutes and 37° for 5 minutes The protein is then added along with 1-10 mM DTT and the incubation continued at 37°C for 5 minutes. Similar heating protocols known in the art may also be used. The protein will then be incubated briefly with the renatured RNA to reconstitute holoenzyme activity.

The RNase P substrates used in the assay are labelled. Examples of labeled nucleotides that can be incorporated into the RNA substrates include BrdUrd (Hoy and Schimke, Mutation Research 290: 217 ,1993), BuUTP (Wansick et al., J. Cell Biology 122:283, 1993) and nucleotides modified with biotin (Langer et al., Proc. Natl. Acad. Sci. USA 78: 6633, 1981) or with suitable haptens such as digoxygenin (Kerhof, Anal. Biochem. 205: 359, 1992). Suitable fluorescence-labeled nucleotides are Fluorescein-isothiocyanate-dUTP, Cyanine-3-dUTP and Cyanine-5-dUTP (Yu et al., Nucleic Acids Res. 22:3226, 1994). A preferred nucleotide analog label for RNA molecules is Biotin-14-cytidine-5'-triphosphate. Fluorescein, Cy3, and Cy5 can be linked to dUTP for direct labeling. Cy3.5 and Cy7 are available as avidin or anti-digoxygenin conjugates for secondary detection of biotin- or digoxygenin-labeled probes.

The amplified rnpA genes may also be cloned into expression vectors ot containing encoded fusion tag sequences, but still containing an inducible promoter. After induction, the overexpressed protein can be purified essentially by the protocol for purification of *E. coli* RNaseP protein (Baer et al., 1990).

Examples of RNA substrates that can be used to measure RNase P enzymatic activity include the full-length substrate $ptRNA^{Tyr}$ (pTyr) (Altman and Kirsebom, *The RNA World*, 2^{nd} Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999), and $ptRNA^{Gln}$ (pGln), an 85-mer from the

cyanobacterium *Synechocystis* (Pascual and Vioque, Proc. Natl. Acad. Sci. USA 96: 6672, 1999) or a substrate obtained from the homologous bacteria.

A modified ptRNA^{Gln} substrate can also be used, in which the 5' end is fluorescently tagged in order to monitor hydrolysis using fluorescence spectroscopy. Given that the chemical synthesis of an 85-mer with a fluorescent tag is technically impractical, and the fluorescent modification enzymatically synthesized RNA is difficult, the preferred method of synthesizing a fluorescently tagged pGln is conducted with the following two steps: a 5' fluorescently modified 26-nucleotide fragment is chemically synthesized and annealed to a 3' 59-nucleotide fragment that has been enzymatically synthesized. These two fragments, when annealed, form a full-length pGln substrate. The unligated junction between the two fragments occurs in the D-loop, a region that is not required for function by the RNase P holoenzyme.

In addition, substrates that contain only the minimally required structural elements for recognition by the enzyme can also be utilized for this reaction, although the Km values for these substrate fragments are usually much higher than the above-described full-length substrates. One example of a substrate fragment is p10AT1, a 45-mer that contains a 10-nucleotide 5' leader sequence, an extended 12-base pair stem which is made up of the aminoacyl acceptor stem, a T-stem, and a single loop. The Km for hydrolysis reactions using this simplified substrate fragment rises to greater than 1μM (McClain et al., 1987). Therefore, while the substrate fragment is easier to constuct, it requires a higher concentration in an enzymatic assay.

The progress of the RNase P-mediated hydrolysis reaction is monitored, for example, by fluorescence spectroscopy. For example, fluorescence polarization assay for RNase P activity is conducted by labeling the 5' end of the substrate, for example, the 45-mer (p10AT1) or the 85-mer (pGln) substrate, with an appropriate fluorophore. Given that compounds in screening libraries often interfere with fluorescence measurements in the blue to yellow region of the spectrum, preferred

fluorophores emit light in the red region of the spectrum (e.g., TAMRA (Molecular Probes, OR) and Cy3 labeled nucleotide (Dharmacon Research, CO.) Samples of the RNase P holoenzyme and the RNase P substrate are mixed, incubated, and measured for spectrophotometric polarization. When the substrate is cleaved by the RNase P holoenzyme, the 10-nucleotide 5'- leader sequence is released, which leads to a substantial change in the fluorescence polarization in the sample. (Campbell, I.D. & Dwed., R.A. pp. 91-125 The Benjamin/Cummings Publishing Company, Menlo Park, CA (1984); Lakowicz, J.R., Plenum Press, NY (1983)).

The preferred reaction buffer contains 50 mM Tris-HCl (pH 7.5), 100 mM ammonium chloride and 10 mM magnesium chloride. Concentrations of 10-100 mM, 25-500 mM and 1-100 mM of the above, respectively, can be substituted, as can other buffering agents such as MOPS or HEPES, or other monovalent cations, such as sodium or potassium. When the assay is run in either 98 or 364-well polystyrene or polypropylene assay plates, there is a very significant decrease in the fluorescence intensity and polarization of the annealed substrate over time in the absence of enzyme. Various conditions have been tested to prevent the loss of signal with time. The preferred conditions include addition of 10-40 μg/ml carbonic anhydrase and 10-100 μg/ml polyC to the buffer. Other materials, such as, 0.5-5% glycerol, 10-100 μg/ml hen egg lysozyme, 10-50 μg/mL tRNA, or 2-10 mM DTT can also be added to the buffer to prevent some loss of signal.

The RNase P hydrolysis rate can also be monitored using a radiolabelled substrate, performing a surface proximity assay (SPA) and measuring hydrolysis by scintillation counting. For example, the substrate is anchored to the surface of the assay plate via a biotin-streptavidin interaction between a biotinylated nucleotide in the anticodon loop and a streptavidin matrix on the plate. The substrate is also ^{33}P-labelled at the 5' end. Using this method, RNase P-mediated hydrolysis of the 5' leader sequence results in decrease scintillation due to reduced proximity of the radiolabel to the scintillation-coated plate. (Brown et al.,

FlashPlate Technology, in J.P. Devlin (Ed.), Marcel Dekker, Inc. NY pp. 317-328.)

A bipartite substrate for RNase P, consisting of a t'-end Cy3 labeled 26mer and *in vitro* T7-polymerase transcribed 59mer is preferred for screening. the 26mer consists of the first 26 contiguous nucleotides of the pre-tRNA substrate including the 10-nucleotide leader sequence. The two RNA fragments are annealed together under appropriate conditions of stoichiometry (59mer in 20 to 100% excess) and temperature in a buffer system consisting of 50 mM Tris-HCl (pH 7.5), 100 mM ammonium chloride and 10 m magnesium chloride. Briefly, the two RNA fragments are mixed together and heated to between 65 and 80°C for about 5 minutes and then slowly cooled to room temperature.

In addition, the RNase P enzyme activity can also be measured using standard techniques described in the literature (see, e.g., Altman and Kirsebom, Ribonuclease P, *The RNA World*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999; Pascual and Vioque, Proc. Natl. Acad. Sci. 96: 6672, 1999; Geurrier-Takada et al., Cell 35: 849, 1983; Tallsjö and Kirsebom, Nucleic Acids Research 21: 51, 1993; Peck-Miller and Altman, J. Mol. Biol. 221: 1, 1991; Gopalan et al., J. Mol. Biol. 267: 818, 1997; and WO 99/11653).

To screen for compounds that inhibit the activity of the RNase P holoenzymes of the present invention, compounds are added to a final concentration of 10 μM before the addition of substrate to the sample. A compound is determined to be an inhibitor if it significantly reduces RNase P hydrolysis as compared to the compound-free control sample. Ideally, the compounds identified as inhibitors selectively inhibit one of the RNase P holoenzymes of the invention without affecting other RNase P holoenzymes. Such inhibitors have the advantage of providing a selective antibacterial treatment that reduces the adverse side effects associated with killing nonpathogenic bacteria. Use of such selective inhibitors also reduces the risk of producing a wide range of resistant bacterial strains.

In general, extracts, compounds, or chemical libraries that can be used in screening assays are known in the art. Examples of such extracts or compounds include, but are not limited to, extracts based on plant, fungal, prokaryotic, or animal sources, fermentation broths, and synthetic compounds, as well as modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (*e.g.*, semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Libraries of genomic DNA or cDNA may be generated by standard techniques (see, *e.g.*, Ausubel *et al., supra*) and are also commercially available (Clontech Laboratories Inc., Palo Alto, CA).

Synthetic compound libraries are commercially available from Brandon Associates (Merrimack, NH) and Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceanographics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are produced, if desired, according to methods known in the art, *e.g.*, by standard extraction and fractionation methods.

When a crude extract is found to modulate an RNase P holoenzyme activity, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible. Thus, the goal of the extraction, fractionation, and purification process is the characterization and identification of a chemical entity within the crude extract having the modulating activities. The same assays described herein for the detection of inhibitors in mixtures of compounds can be used to purify the active component and to test derivatives thereof. Methods of fractionation and purification of such heterogenous extracts are known in the art. If desired, compounds shown to be useful agents for treatment are chemically modified according to methods known in the art.

Compounds which modulate an RNase P holoenzyme activity may be administered by any appropriate route for treatment or prevention of a disease or condition associated a bacterial infection. Administration may be topical, parenteral, intravenous, intra-arterial, subcutaneous, intramuscular, intracranial, intraorbital, ophthalmic, intraventricular, intracapsular, intraspinal, intracisternal, intraperitoneal, intranasal, aerosol, by suppositories, or oral administration.

Therapeutic formulations may be in the form of liquid solutions or suspensions; for oral administration, formulations may be in the form of tablets or capsules; and for intranasal formulations, in the form of powders, nasal drops, or aerosols.

Methods well known in the art for making formulations are found, for example, in "Remington's Pharmaceutical Sciences." Formulations for parenteral administration may, for example, contain excipients, sterile water, or saline, polyalkylene glycols such as polyethylene glycol, oils of vegetable origin, or hydrogenated napthalenes. Biocompatible, biodegradable lactide polymer, lactide/glycolide copolymer, or polyoxyethylene-polyoxypropylene copolymers may be used to control the release of the compounds. Other potentially useful parenteral delivery systems include ethylene-vinyl acetate copolymer particles, osmotic pumps, implantable infusion systems, and liposomes. Formulations for inhalation may contain excipients, for example, lactose, or may be aqueous solutions containing, for example, polyoxyethylene-9-lauryl ether, glycholate and deoxycholate, or may be oily solutions for administration in the form of nasal drops, or as a gel. The concentration of the compound in the formulation will vary depending upon a number of factors, including the dosage of the drug to be administered, and the route of administration.

The formulations can be administered to human patients in therapeutically effective amounts (e.g., amounts which prevent, eliminate, or reduce a pathological condition) to provide therapy for a disease or condition associated with infection. Typical dose ranges are from about 0.1 μg/kg to about 1 g/kg of

body weight per day. The preferred dosage of drug to be administered is likely to depend on such variables as the type and extent of the disorder, the overall health status of the particular patient, the formulation of the compound excipients, and its route of administration.

Other Embodiments

All publications and patent applications mentioned in this specification are herein incorporated by reference.

While the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications. Therefore, this application is intended to cover any variations, uses, or adaptations of the invention that follow, in general, the principles of the invention, including departures from the present disclosure that come within known or customary practice within the art. Other embodiments are within the claims.

What is claimed is:

1. An isolated polypeptide comprising an RNase P consensus sequence wherein said polypeptide has RNase P protein activity.

2. The polypeptide of claim 1, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOS: 20-38.

3. An isolated nucleic acid sequence, wherein said sequence encodes a polypeptide containing an RNase P consensus and said polypeptide has RNase P protein activity.

4. The nucleic acid sequence of claim 3, wherein said sequence encodes a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 20-38.

5. The nucleic acid sequence of claim 4, wherein said sequence is selected from the group consisting of SEQ ID NOS: 1-19.

6. A transgenic host cell, wherein said cell comprises a heterologous nucleic acid sequence encoding the polypeptide of claim 1.

7. An antibody that specifically binds to the polypeptide of claim 1.

8. A method of identifying an antibiotic agent, said method comprising:

i) obtaining an RNase P holoenzyme comprising the polypeptide of claim 1;

ii) contacting said holoenzyme with an RNase P substrate in the presence and in the absence of a compound; and

iii) measuring the enzymatic activity of said holoenzyme;

wherein a compound is identified as an antibiotic agent if said compound produces a detectable decrease in said RNase P enzymatic activity as compared to activity in the absence of said compound.

9. The method of claim 8, wherein said polypeptide is substantially identical to a polypeptide of SEQ ID NOS:20-38.

10. The method of claim 8, wherein said activity is measured by fluorescence spectroscopy.

11. The method of claim 8, wherein said RNase substrate is fluorescently tagged $ptRNA^{Gln}$.

12. A method for making a $ptRNA^{Gln}$, said method comprising annealing two RNA fragments together by heating to about 65°C to about 80°C for about 5 minutes, followed by cooling to 20-25° C.

13. The method of claim 8, wherein said fluorescence analysis is carried out in a buffer comprising 10-40 mg/ml carbonic anhydrase and 10-100 μg/ml polyC.

14. The method of claim 13, wherein said buffer further comprises at least one of the following:

0.5-5% glycerol;

10-100 μg/ml hen egg lysozyme;

10-50 μg/ml tRNA; or

1-10 mM DTT.

NOVEL BACTERIAL RNase P PROTEINS AND THEIR USE IN IDENTIFYING ANTIBACTERIAL COMPOUNDS

Abstract of the Disclosure

The invention features novel RNase P molecules and nucleic acids encoding the same. Methods for discovery of antimicrobial compounds are also featured.

50093.016001 RNase P application.wpd

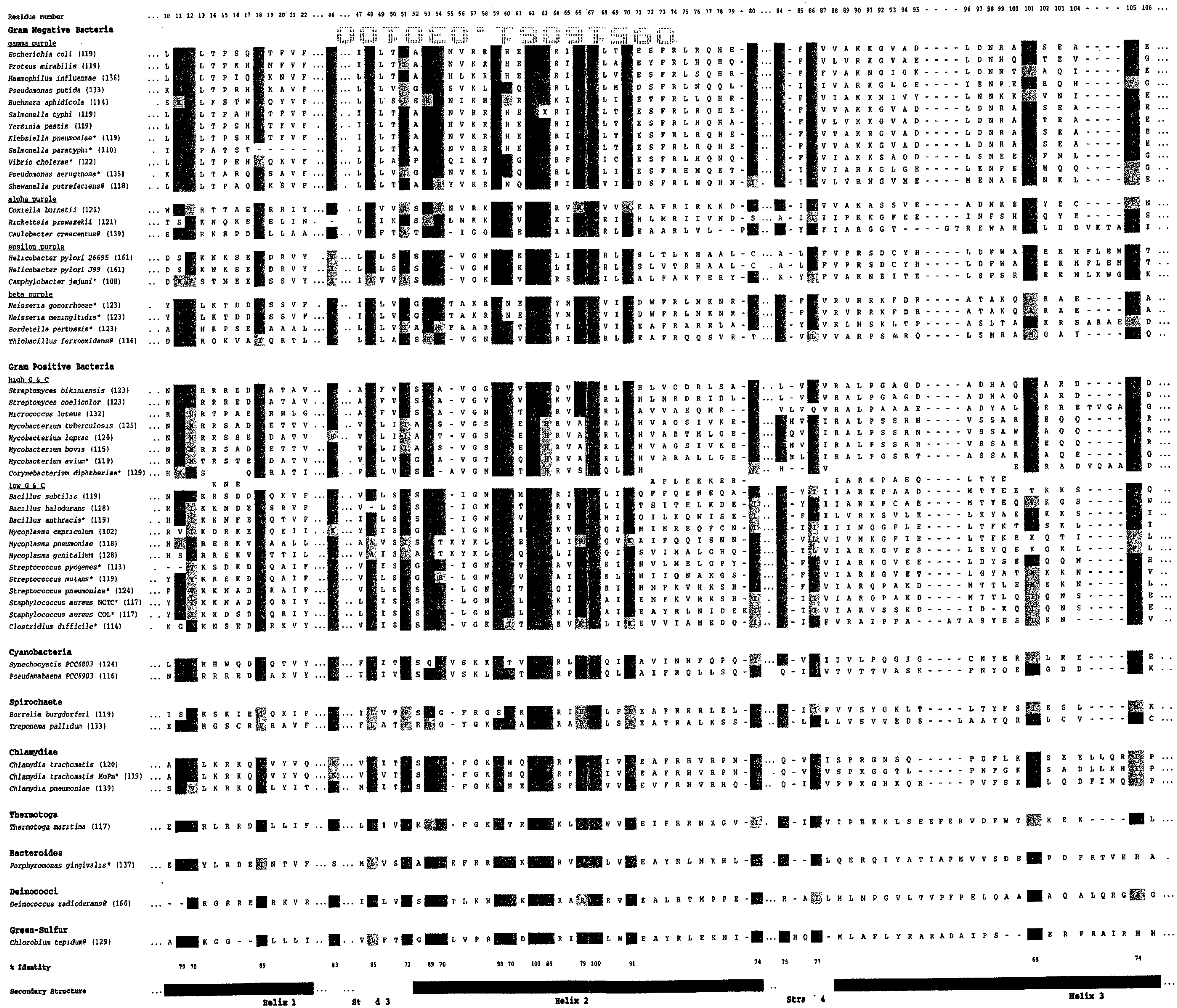

FIG. 1

***Streptococcus mutans* UAB159** (119 aa)
Amino acid sequence:
VLKKAYRVKSDKDFQAIFTEGRSVANRKFVVYSLEKDQSHYRVGLSVGKRLGNAVVRNAIKRKLRHVLMELGPYLGT
QDFVVIARKGVEELDYSTMKKNLVHVLKLAKLYQEGSIREKE
Nucleotide sequence (plus strand):
AGATTTTTGGCTTTTTCTCATTTTATGATATAATAGTGATAATTTAAATATTGGAGTCAT***GTT***TTGAAAAAAGCCTA
TCGCGTTAAAAGTGATAAAGATTTTCAGGCAATTTTTACTGAAGGACGAAGTGTTGCCAATCGGAAATTTGTTGTCT
ATAGTTTAGAAAAAGATCAAAGTCACTATCGTGTTGGACTTTCAGTTGGAAAAAGATTAGGAAATGCTGTCGTTAGA
AATGCGATTAAACGAAAATTGCGCCATGTCCTTATGGAACTTGGTCCTTATTTAGGCACTCAAGATTTTGTTGTTAT
TGCTAGAAAAGGTGTTGAGGAACTTGATTATAGCACGATGAAAAAAAATCTGGTTCATGTTTTAAAACTGGCTAAAC
TGTATCAGGAAGGATCTATTCGTGAAAAAGAA
Sequence origin: University of Oklahoma ACGT; Contig 299

FIG. 2A

***Klebsiella pneumoniae* M6H 78578** (119 aa)
Amino acid sequence:
VVKLAFPRELRLLTPSHFTFVFQQPQRAGTPQITILGRLNSLGHPRIGLTVAKKNVKRAHERNRIKRLTRESFRLRQ
HELPPMDFVVVAKRGVADLDNRALSEALEKLWRRHCRLARGS
Nucleotide sequence (plus strand):
CGTCGTCGTGCTAAAGGCCGCGCTCGTCTGACCGTTTCCAAGTAATAAAGCTAACCCTGC***GTG***GTTAAGCTCGCATT
TCCCAGGGAGTTACGCTTGTTAACTCCCAGTCATTTCACTTTCGTCTTCCAGCAGCCACAACGGGCTGGCACGCCGC
AAATCACCATCCTCGGCCGCCTGAATTCGCTGGGGCATCCCCGCATCGGTCTCACCGTCGCCAAGAAAAACGTGAAA
CGCGCACATGAACGCAATCGGATTAAACGTCTGACGCGTGAAAGTTTTCGTTTGCGTCAACATGAACTCCCGCCAAT
GGATTTCGTGGTGGTGGCGAAAAGAGGGGTTGCCGACCTCGATAACCGTGCTCTCTCGGAAGCGTTGGAAAAATTAT
GGCGCCGCCATTGTCGCCTGGCTCGCGGGTCC**TGA**TCGGCCTGATTCGAGTTTATCAGCGCCTGATTAGTCCGCTAC
TCGGGCCGCATTGTC
Sequence origin: Washington University; Contig 632

FIG. 2B

***Salmonella paratyphi* A ATCC 9150** (110 aa)
Amino acid sequence:
VTFVNSRSFHIRLPATSTGCTPQITILGRLNSLGHPRIGLTVAKKNVRRAHERNRIKRLTRESFRLRQHELPAMDFV
VVAKKGVADLDNRALSEALEKLWRRHCRLARGS
Nucleotide sequence (plus strand):
CTGACCGTTTCCAAGTAATAAAGCTAACCCCTGAGTGGTTAAGCTCGCATTTCCCAGGGA***GTT***ACGTTTGTTAACTC
CCGCTCATTTCACATTCGTCTTCCAGCAACCTCAACGGGCTGCACGCCGCAAATCACCATCCTCGGCCGCCTGAATT
CGCTGGGGCATCCCCGTATCGGTCTTACCGTCGCCAAGAAAAATGTTCGACGTGCGCATGAACGCAACCGGATTAAA
CGTCTGACGCGTGAAAGCTTCCGTCTGCGCCAGCATGAACTTCCTGCAATGGATTTCGTGGTGGTGGCGAAAAAAGG
GGTTGCCGACCTCGATAACCGTGCTCTCTCGGAAGCGTTGGAAAAATTATGGCGCCGCCACTGTCGCCTGGCTCGCG
GGTCCTGA**TAG**CCCTTATTCGGGTCTATCAACGCCTGATCAGTCCGCTGCTTGGGCCGCATTGTCGTTTC
Sequence origin: Washington University;

FIG. 2C

***Pseudomonas aeruginosa* PAO1** (135 aa)
Amino acid sequence:
VVSRDFDRDKRLLTARQFSAVFDSPTGKVPGKHVLLLARENGLDHPRLGLVIGKKNVKLAVQRNRLKRLIRESFRHN
QETLAGWDIVVIARKGLGELENPELHQQFGKLWKRLLRNRPRTESPADAPGVADGTHA
Nucleotide sequence (plus strand):
TCTGTCGCGTCGTCGCGCCAAAGGCCGTAAGCGTCTGACCGTCTGATTTATCCGGTACGG***GTG***GTGAGTCGGGACTT
CGACCGGGACAAGCGTCTACTGACAGCCCGGCAATTCAGCGCAGTCTTCGACTCTCCGACCGGCAAGGTCCCCGGCA
AGCACGTCCTGCTGCTGGCGCGCGAGAACGGTCTCGATCACCCCCGCCTGGGCCTGGTGATCGGCAAGAAGAACGTC
AAGCTCGCCGTCCAGCGCAATCGCCTCAAACGCCTGATCCGCGAATCGTTCCGCCATAACCAGGAAACCCTGGCTGG
CTGGGATATCGTGGTGATCGCGCGCAAAGGCCTGGGCGAACTGGAAAATCCGGAGCTGCACCAGCAGTTCGGCAAGC
TCTGGAAACGCCTGTTGCGCAATCGACCTCGCACGGAAAGCCCTGCTGACGCCCCTGGCGTGGCCGACGGTACTCAT
GCA**TAG**GTCGATGCCCGCGCATCCCGATCCCTGTAGTGTCATCCCCCCTTCGATGACCCGGCACCG
Sequence origin: Pathogenesis & University of Washington; Contig 54

FIG. 2D

Corynebacterium diphtheriae (129 aa)
Amino acid sequence:
VTLTSSNRTTVLPSQHKLSNSEQFRATIRKGKRAGRSTVVLHFYAEATAGNLATAGGPRFGLVVSKAVGNAVTRHRV
SRQLRHVVIAMKDQFPASSHVVVRAIPPAATASYEELRADVQAALDKLNRKR
Nucleotide sequence (plus strand):
CCGGTCGCGCAATCGTGGCTGCACGTCGTAACAAGGGTCGTAAGAGCCTGACCGCTTAAG***GTC***ACTCTTACAAGCTC
GAATAGAACGACGGTGCTACCTTCACAGCACAAGCTCAGCAATTCCGAACAGTTCCGCGCAACGATTCGGAAGGGCA
AGCGTGCTGGGAGGAGCACCGTCGTTCTTCATTTTTATGCTGAGGCGACCGCGGGCAACCTTGCAACCGCAGGCGGC
CCGCGATTCGGCCTCGTTGTGTCCAAGGCTGTTGGAAATGCTGTGACTCGTCACCGTGTTTCGCGGCAGTTAAGGCA
CGTAGTAATCGCTATGAAAGACCAGTTCCCAGCGTCATCCCATGTTGTTGTGAGGGCGATACCGCCAGCGGCGACAG
CAAGTTATGAGGAGTTGCGGGCAGATGTGCAGGCAGCACTCGACAAGCTCAACCGCAAGCGA**TAA**GGCGGTTACTCG
CCCTCGTGGGCTGGTTAGTCGCGCATTGTTTGATGCGGTGCGGTTCTA
Sequence origin: Sanger centre; Contig 390

FIG. 2E

***Chlamydia trachomatis* MoPn** (119 aa)
Amino acid sequence:
VHRLTLPKSARLLKRKQFVYVQRCGQYCRTDQATLRIVPSRHSNIRKVGVTVSKKFGKAHQRNRFKRIVREAFRHVR
PNLPACQVVVSPKGGTLPNFGKLSADLLKHIPEALPLVTSSK
Nucleotide sequence (plus strand):
GCTACAAAAAGTGGAAGAAATCTTTTAAATCGTCGTCGCCGTCACGGCAGACATTCCTTAATTGATCTCTAAGATCT
TTCATTT***GTG***CATCGGTTAACTCTACCTAAAAGTGCCCGCCTATTGAAACGTAAACAATTTGTTTACGTGCAGCGTT
GTGGGCAATATTGTCGTACTGATCAGGCAACTTTACGAATAGTTCCTTCTCGTCATTCGAACATCCGTAAAGTAGGG
GTTACTGTTTCTAAAAAATTTGGGAAAGCCCATCAGCGCAATCGCTTTAAAAGAATTGTGCGAGAGGCTTTTAGGCA
TGTGCGACCAAATCTTCCCGCATGTCAAGTGGTAGTGTCTCCTAAAGGGGGCACTCTACCAAATTTTGGTAAACTAT
CCGCGGATCTTCTTAAGCATATTCCAGAGGCTTTGCCTCTCGTTACTTCTTCTAAG**TAG**TTTTTTATTTTGGTCAAA
AAATAAAAAAACCATTCCACGCTATAGAGGCATGGAATGGGAA
Sequence origin: TIGR & Manitoba University;

FIG. 2F

***Vibrio cholerae* serotype O1, Biotype El Tor, Strain N16961** (122 aa)
Amino acid sequence:
SRIILSTYAFNRELRLLTPEHYQKVFQQAHSAGSPHLTIIARANNLSHPRLGLAVPKKQIKTAVGRNRFKRICRESF
RLHQNQLANKDFVVIAKKSAQDLSNEELFNLLGKLWQRLSRPSRG
Nucleotide sequence (minus strand): *NO INITIATOR CODON BEFORE STOP*
GGCAGCGTGGGCCGATAAGTGGACTAATAAACCACTGGTAAAGTTTACAATACCAATGG**CTA**ACCACGAGAAGGGC
GAGAGAGGCGTTGCCATAGTTTGCCAAGCAAGTTAAACAGTTCTTCATTGCTCAAATCTTGCGCGCTCTTTTTGGCG
ATGACAACAAAATCTTTGTTAGCCAGTTGATTTTGATGTAAGCGAAAGCTTTCTCTGCAAATACGTTTGAATCGATT
ACGGCCGACGGCAGTTTTGATCTGCTTTTTAGGAACCGCGAGTCCCAAACGAGGATGAGAAAGGTTATTAGCGCGAG
CGATGATTGTGAGATGAGGAGAACCAGCACTGTGAGCTTGCTGGAAGACTTTTTGATAATGTTCGGGAGTTAACAAA
CGTAACTCCCGATTGAATGCGTACGTACTCAAAATAATTCGAGA**TTA**TTTGACAGGCGCTTACGGCCTTTTGCACG
ACGTGCATTCAGAACTTTACGACCGTTCGC
Sequence origin: TIGR

FIG. 2G

***Neisseria gonorrhoea* FA 1090** (123 aa)
Amino acid sequence:
VILDYRFGRQYRLLKTDDFSSVFAFRNRRSRDLLQVSRSNGNGLDHPRIGLVVGKKTAKRANERNYMKRVIRDWFRL
NKNRLPPQDFVVRVRRKFDRATAKQARAELAQLMFGNPATGCGKQV
Nucleotide sequence (minus strand):
ATGTTCCTTGTATGGGAAACCCGTTGCCGTCTGAACCTTGCCTGCAGGGTACCGTTCTGA**TCA**TACCTGTTTCCCGC
ATCCGGTTGCGGGGTTGCCGAACATGAGTTGTGCCAGTTCCGCCCTTGCCTGTTTTGCGGTAGCCCTGTCGAATTTC
CGGCGGACGCGCACGACGAAATCCTGAGGCGGCAGCCGGTTTTTGTTCAATCTGAACCAGTCGCGGATGACGCGTTT
CATATAGTTCCGCTCGTTGGCGCGTTTGGCGGTTTTTTTGCCGACCACCAGACCGATGCGGGGATGGTCCAGCCCGT
TGCCGTTTGAGCGCGAAACTTGCAGCAGGTCGCGGCTGCGGCGGTTTCTGAATGCAAAAACGGATGAAAAATCATCC
GTTTTTAACAAGCGGTACTGCCTTCCGAAGCGGTAGTCCAAAAT***TAC***ACTGCCAGGCGTTTGCGGCCTTTGGCACGG
CGTGCGGCCAATACTGCGCGTCCGCCGCGT
Sequence origin: University of Oklahoma ACGT; Contig 60

FIG. 2H

***Neisseria meningitidis* serogroup A Strain Z2491** (123 aa)
Amino acid sequence:
VILDYRFGRQYRLLKTDDFSSVFAFRNRRSRDLLQVSRSNGNGLDHPRIGLVVGKKTAKRANERNYMKRVIRDWFRL
NKNRLPPQDFVVRVRRKFDRATAKQARAELAQLMFGNPATGCRKQA
Nucleotide sequence (minus strand):
TGTTCCTTAGTATGGGAAACCCGTTGCCGTCTGAACCTTGCCTGCAGAGTACCGTTCTGA**TCA**TGCCTGTTTCCTGC
ATCCGGTTGCGGGGTTGCCGAACATGAGTTGTGCCAGTTCCGCCCTTGCCTGTTTTGCGGTAGCCCTGTCGAATTTA
CGGCGGACGCGCACGACGAAATCCTGCGGCGGCAGCCGGTTTTTGTTCAATCTGAACCAGTCGCGGATGACGCGCTT
CATATAATTTCGTTCGTTGGCGCGTTTGGCGGTTTTTTTGCCGACCACCAGACCGATGCGGGGATGATCCAGCCCGT
TGCCGTTTGAACGCGAAACTTGCAGCAGGTCGCGGCTGCGGCGGTTTCTGAATGCAAAAACGGATGAAAAATCATCC
GTTTTCAACAAGCGGTACTGCCTTCCGAAGCGGTAGTCCAAAAT***TAC***ACCGCCAGGCGTTTGCGGCCTTTGGCGCGC
CGTGCGGCCAATACTGCGCGTCCGCCGCGC
Sequence origin: Sanger centre & Oxford University

FIG. 2I

***Streptococcus pyogenes* M1** (113 aa)
Amino acid sequence:
VKREKDFQAIFKDGKSTANRKFVIYHLNRGQDHFRVGISVGKKIGNAVTRNAVKRKIRHVIMALGHQLKSEDFVVIA
RKGVESLEYQELQQNLHHVLKLAQLLEKGFESEEKH
Nucleotide sequence (minus strand):
GTTACCTCACCACGACCACAGGCCACTAATAATAGAACTAAGGGGACTATTCTTGCAATT**TTA**ATGTTTTTCTTCAC
TCTCAAAACCTTTCTCAAGCAATTGTGCTAACTTTAAAACATGATGTAAATTTTGTTGAAGCTCTTGATACTCCAAA
GATTCGACACCCTTACGGGCAATCACCACGAAATCCTCTGACTTCAGCTGATGCCCTAATGCCATGATAACATGACG
TATCTTTCGTTTGACTGCATTTCTGGTGACTGCATTTCCTATTTTTTACCGACAGAAATACCCACACGGAAGTGGT
CTTGGCCTCTATTTAAATGATAAATGACAAATTTTCGATTTGCTGTACTTTTTCCATCCTTAAATATGGCTTGGAAA
TCTTTCTCACGCTT***GAC***ACGATAGGTCTTCTTCAAAATTTAACTCCAATATCTAAATTATTACCATTATACCACATC
Sequence origin: University of Oklahoma ACGT; Contig 7

FIG. 2J

***Bordetella pertussis* Tohama I** (123 aa)
Amino acid sequence:
MPRATLPAEARLHRPSEFAAALKGRRLARGAFFIVSASPCAPADDQPARARLGLVIAKRFAARAVTRNTLKRVIREA
FRARRLALPAQDYVVRLHSKLTPASLTALKRSARAEVDAHFTRIAR
Nucleotide sequence (minus strand):
CCACCCAGGGGCTGAGGAAGTACCGGTAAAACCGGATCGGGGCGATAAGCAGTCTCCTGA**TCA**TCGCGCTATCCGTG
TGAAGTGAGCATCTACTTCGGCGCGCGCCGAGCGTTTCAGGGCCGTGAGGCTTGCCGGTGTCAGCTTGCTGTGCAGC
CGCACCACGTAATCCTGGGCCGGCAGGGCAAGCCGGCGAGCCCGGAACGCTTCGCGGATGACCCGCTTCAAGGTATT
GCGCGTCACGGCGCGGGCGGCAAAACGCTTGGCGATCACCAGGCCCAGGCGCGCGCGCGCCGGCTGGTCATCAGCAG
GGGCACAGGGCGAGGCGCTGACAATAAAGAAAGCCCCTCGGGCCAGTCGCCGGCCTTTGAGGGCGGCGGCAAACTCG
GAGGGGCGATGCAATCGCGCCTCCGCAGGGAGCGTGGCGCGCGG***CAT***GGGTGACGTGACGGAGACTGGCGACGGGGC
CGGCGGCGATGCTCCTGTTACAGGCAATCC
Sequence origin: Sanger centre & MDS; Contig 267

FIG. 2K

***Porphyromonas gingivalis* W83** (137 aa)
Amino acid sequence:
MTSPPTFGLSKSERLYLRDEINTVFGEGKAFVVYPLRVVYRLGSEHRVAYSSMLVSVAKKRFRRAVKRNRVKRLVRE
AYRLNKHLLNDVLQERQIYATIAFMVVSDELPDFRTVERAMQKSLIRIAGNVPSSALKNE
Nucleotide sequence (minus strand):
AGAAGAAAATGGGGAGCAGTAAGAGTTGCACGAGAAAAGCCTTGATCAGTCGCATCGTAT**TTA**CTCGTTTTTCAAAG
CCGATGAAGGTACATTTCCGGCAATTCTGATCAGACTCTTTTGCATCGCTCTCTCCACTGTACGAAAGTCAGGAAGT
TCATCCGATACTACCATAAATGCAATAGTAGCATAGATCTGTCTCTCTTGGAGGACATCGTTCAGGAGGTGTTTGTT
GAGCCGATAAGCCTCCCTGACCAAACGCTTGACCCTATTGCGCTTCACGGCTCGCCTAAACCTTTTCTTTGCTACGC
TTACCAGCATGGAGGAATATGCAACTCGATGCTCCGATCCCAGACGGTAGACTACGCGTAGAGGATAAACGACAAAC
GCCTTGCCTTCGCCAAAGACCGTATTGATTTCATCGCGAAGATAGAGGCGTTCGCTTTTGGATAGGCCGAATGTAGG
CGGAGAGGT***CAT***TTCCCGTTGAGGTAATCCTCTAATGCCATAGCCATAGAAGGATATTGCTCGGTCGGCGCA
Sequence origin: TIGR & Forsyth Dental Center

FIG. 2L

***Streptococcus pneumoniae* Type 4** (124 aa)
Amino acid sequence:
VLKKNFRVKREKDFKAIFKEGTSFANRKFVVYQLENQKNRFRVGLSVSKKLGNAVTRNQIKRRIRHIIQNAKGSLVE
DVDFVVIARKGVETLGYAEMEKNLLHVLKLSKIYREGNGSEKETKVD
Nucleotide sequence (minus strand):
TCGCTAGTTACCCCATTAGTCGCACAGGCTGTCATGATTAACAGAGACAGTCCTAGCAAA**CTA**GTCAACTTTAGTTT
CTTTTTCACTCCCATTTCCTTCCCGGTAAATCTTTGATAATTTTAATACATGGAGTAGATTTTTCTCCATCTCTGCG
TATCCCAAGGTTTCGACTCCTTTTCGAGCAATGACAACAAAGTCGACATCTTCTACCAGACTCCCTTTTGCATTCTG
GATAATATGCCGAATCCGTCGCTTAATTTGATTTCTAGTGACGGCATTCCCCAGTTTTTTGCTAACTGATAGACCTA
CTCGAAAACGGTTTTTCTGGTTTTCTAATTGGTAGACCACAAATTTGCGATTAGCAAAACTTGTCCCCTCCTTGAAA
ATCGCCTTAAAATCTTTCTCTCTTTTTACACGAAAGTTTTTCTTCAA***AAC***TCAACTCCATCTATTAAATTACTACTA
TTATACCATATTTTTCAAAAAAGCCAATCATAG
Sequence origin: TIGR;

FIG. 2M

***Clostridium difficile* 630 (epidemic type X)** (114 aa)
Amino acid sequence:
MDFNRTKGLKKDSDFRKVYKHGKSFANKYLVIYILKNKSDYSRVGISVSKKVGKAITRNRVRRLIKEAYRLNIDEKI
KPGYDIVFIARVSSKDATFKDIDKSIKNLVKRTDISI
Nucleotide sequence (minus strand):
TCCTTTAATATATAAATTATTTTATTCAAAGTCATTAACCTCCATATTTATAGCATACAA**TTA**AATAGAAATATCCG
TTCTTTTAACTAAATTTTTTATAGACTTGTCTATGTCTTTAAAAGTAGCATCCTTACTAGATACCCTTGCTATAAAT
ACTATATCATATCCAGGCTTAATTTTTTCATCAATATTTAATCTGTAGGCTTCTTTTATTAATCTTCTTACTCTATT
CCTAGTAATAGCTTTTCCTACTTTTTTTGAAACAGAAATACCTACTCTACTATAATCTGATTTATTTTTAAGTATAT
ATATTACTAAATATTTGTTTGCAAAAGATTTGCCGTGTTTATATACTTTTCTAAAATCAGAGTCTTTTTTCAACCCT
TTAGTCCTATTAAAGTC***CAT***AGTTAACCTCCATAAACACAGCTATGAATCGTAATTATTTACACAAAAAGGCCACCT
TTG
Sequence origin: Sanger centre; Contig 975

FIG. 2N

***Camphylobacter jejuni* NCTC** (108 aa)
Amino acid sequence:
VKNFDKFSTNEEFSSVYKVGKKWHCEGVIIFYLNSYEKKIAVVASKKVGKAVVRNRSKRILRALFAKFERYLQDGKY
IFVAKNEITELSFSRLEKNLKWGLKKLECFK
Nucleotide sequence (minus strand):
AAGCAGCGGGTTTTAAAGGGCTTAAGAATTTCTGATAAAAACGGAGTATTTTTAGGCATA**TCA**TTTGAAACATTCTA
GTTTTTTCAATCCCCATTTTAGATTTTTTTCTAACCTAGAAAAAGAAAGTTCAGTGATTTCATTTTTAGCTACAAAA
ATATATTTGCCATCTTGAAGATATCTTTCAAACTTAGCAAACAAAGCTCTTAAAATTCGTTTTGAACGATTTCTAAC
CACTGCTTTTCCAACTTTTTTACTAGCAACAACTGCTATTTTTTTTTCATAACTATTCAGATAAAAAATGATCACAC
CTTCGCAATGCCATTTTTTGCCTACTTTATATACAGATGAAAATTCCTCGTTTGTGCTAAATTTATCAAAATTTTT***C***
ACACAGCAAGTCTTTTTCTACCTTTAGCGCGTCTTGCATTGATCACTTTGCGACCATTTTTA
Sequence origin: Sanger centre & MDS

FIG. 2O

***Bacillus anthracis* Ames** (119 aa)
Amino acid sequence:
MKKKHRIKKNDEFQTVFQKGKSNANRQFVVYQLDKEEQPNFRIGLSVSKKIGNAVVRNRIKRMIRQSITELKDEIDS
GKDFVIIARKPCAEMTYEELKKSLIHVFKRSGMKRIKSSVRK
Nucleotide sequence (minus strand):
TAAACCTAATTTCTTTTTCAAAGCCTACTCCTCCTTGTATCGGTATGTATATAGTGTAAT**TCA**TTTCCTTACGCTAC
TTTTTATTCTTTTCATACCAGAGCGTTTAAAGACATGAATTAAGCTTTTCTTTAATTCTTCATATGTCATCTCTGCA
CAAGGCTTCCTTGCTATTATAACAAAATCTTTTCCAGAATCTATCTCATCTTTTAATTCTGTGATCGACTGGCGAAT
CATACGTTTAATTCGGTTACGCACTACTGCATTTCCTATCTTCTTGCTGACAGAAAGGCCAATACGAAAGTTTGGCT
GCTCTTCTTTATCTAGTTGATAGACAACAAATTGACGATTCGCATTCGATTTTCCTTTTTGAAAAACCGTCTGGAAT
TCATCATTCTTTTTTATACGATGTTTTTTCTT***CAT***ATCAATTGACACTCCTGTAGTTCATCAGCGGAAATTCACTAT
TATTAGAAAAAAAGACCA
Sequence origin: TIGR;

FIG. 2P

***Mycobacterium avium* 104** (119 aa)
Amino acid sequence:
VLPARNRMTRSTEFDATVKHGTRMAQPDIVVHLRRDSEPDDESAGPRVGLVVGKAVGTAVQRHRVARRLRHVARALL
GELEPSDRLVIRALPGSRTASSARLAQELQRCLRRMPAGTGP
Nucleotide sequence (minus strand):
GTCCGCGGGCGACGGTTCGGCCGGCGCCGCGAATGGCCGCGCCCGACCGCGCCGGTCCGG**TCA**CGGCCCGGTTCCCG
CCGGCATGCGCCGCAGGCACCGCTGCAGTTCCTGCGCCAGGCGCGCCGACGACGCGGTCCGGCTTCCGGGCAGCGCG
CGAATCACCAGCCGGTCGGATGGTTCGAGTTCGCCGAGCAGGGCCCGGGCCACGTGACGCAGCCGGCGGGCCACGCG
GTGTCGTTGCACCGCCGTCCCGACGGCCTTCCCGACGACCAGCCCGACCCGTGGGCCCGCGGATTCGTCGTCGGGTT
CGGAGTCGCGCCGGAGGTGGACGACGATGTCGGGCTGCGCCATGCGGGTTCCGTGCTTCACCGTCGCGTCAAACTCG
GTTGACCGCGTCATGCGGTTGCGTGCGGGAAG***CAC***CGCGAAAGACCTGACGTGCGATCAGGCAGAGAGCGCGCGGCG
ACCCTTGCGGCGCCGACC
Sequence origin: TIGR;

FIG. 2Q

***Staphylococcus aureus* NCTC 8325** (117 aa)
Amino acid sequence:
MLLEKAYRIKKNADFQRIYKKGHSVANRQFVVYTCNNKEIDHFRLGISVSKKLGNAVLRNKIKRAIRENFKVHKSHI
LAKDIIVIARQPAKDMTTLQIQNSLEHVLKIAKVFNKKIK
Nucleotide sequence (plus strand):
GTTATAAGCTCAATAGAAGTTTAAATATAGCTTCAAATAAAAACGATAAATAAGCGAGTG***ATG***TTATTGGAAAAAGC
TTACCGAATTAAAAAGAATGCAGATTTTCAGAGAATATATAAAAAAGGTCATTCTGTAGCCAACAGACAATTTGTTG
TATACACTTGTAATAATAAAGAAATAGACCATTTTCGCTTAGGTATTAGTGTTTCTAAAAAACTAGGTAATGCAGTG
TTAAGAAACAAGATTAAAAGAGCAATACGTGAAAATTTCAAAGTACATAAGTCGCATATATTGGCCAAAGATATTAT
TGTAATAGCAAGACAGCCAGCTAAAGATATGACGACTTTACAAATACAGAATAGTCTTGAGCACGTACTTAAAATTG
CCAAAGTTTTTAATAAAAAGATTAAG**TAA**GGATAGGGTAGGGGAAGGAAAACATTAACCACTCAACACATCCCGAAG
TCTTACCTCAGA
Sequence origin: University of Oklahoma ACGT; Contig 561

FIG. 2R

***Staphylococcus aureus* COL** (117 aa)
Amino acid sequence:
MLLEKAYRIKKNADFQRIYKKGHSVANRQFVVYTCNNKEIDHFRLGISVSKKLGNAVLRNKIKRAIRENFKVHKSHI
LAKDIIVIARQPAKDMTTLQIQNSLEHVLKIAKVFNKKIK
Nucleotide sequence (plus strand):
GTTATAAGCTCAATAGAAGTTTAAATATAGCTTCAAATAAAAACGATAAATAAGCGAGTG***ATG***TTATTGGAAAAAGC
TTACCGAATTAAAAAGAATGCAGATTTTCAGAGAATATATAAAAAAGGTCATTCTGTAGCCAACAGACAATTTGTTG
TATACACTTGTAATAATAAAGAAATAGACCATTTTCGCTTAGGTATTAGTGTTTCTAAAAAACTAGGTAATGCAGTG
TTAAGAAACAAGATTAAAAGAGCAATACGTGAAAATTTCAAAGTACATAAGTCGCATATATTGGCCAAAGATATTAT
TGTAATAGCAAGACAGCCAGCTAAAGATATGACGACTTTACAAATACAGAATAGTCTTGAGCACGTACTTAAAATTG
CCAAAGTTTTTAATAAAAAGATTAAG**TAA**GGATAGGGTAGGGGAAGGAAAACATTAACCACTCAACACATCCCGAAG
TCTTACCTCAGA
Sequence origin: TIGR;

FIG. 2S

PATENT
ATTORNEY DOCKET NO: 50093/016001

COMBINED DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am joint inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled NOVEL BACTERIAL RNASE P PROTEINS AND THEIR USE IN IDENTIFYING ANTIBACTERIAL COMPOUNDS, the specification of which

- ☒ is attached hereto.
- ☐ was filed on ________________ as Application Serial No. ________________
  and was amended on ________________.
- ☐ was described and claimed in PCT International Application No. ________________
  filed on ____________ and as amended under PCT Article 19 on ____________.

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose all information I know to be material to patentability in accordance with Title 37, Code of Federal Regulations, §1.56(a).

FOREIGN PRIORITY RIGHTS: I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America filed by me on the same subject matter having a filing date before that of the application(s) of which priority is claimed:

Country	Serial Number	Filing Date	Priority Claimed?
			Yes/No

PROVISIONAL PRIORITY RIGHTS: I hereby claim priority benefits under Title 35, United States Code, §119(e) and §120 of any United States provisional patent application(s) listed below filed by an inventor or inventors on the same subject matter as the present application and having a filing date before that of the application(s) of which priority is claimed:

Serial Number	Filing Date	Status

NON-PROVISIONAL PRIORITY RIGHTS: I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose all information

I know to be material to patentability as defined in Title 37, Code of Federal Regulations, §1.56(a) which became available between the filing date of the prior application and the national or PCT international filing date of this application:

Serial Number	Filing Date	Status

I hereby appoint the following attorneys and/or agents to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith: Paul T. Clark, Reg. No. 30,162, Karen L. Elbing, Ph.D. Reg. No. 35,238, Kristina Bieker-Brady, Ph.D. Reg. No. 39,109, Susan M. Michaud, Ph.D. Reg. No. 42,885, Mary Rose Scozzafava, Ph.D., Reg. No.36,268, James D. DeCamp, Ph.D., Reg. No. 43,580..

Address all telephone calls to: Kristina Bieker-Brady, Ph.D. at 617/428-0200.

Address all correspondence to: Kristina Bieker-Brady, Ph.D. at Clark & Elbing LLP, 176 Federal Street, Boston, MA 02110.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patents issued thereon.

Full Name (First, Middle, Last)	Residence Address (City, State, Country)	Post Office Address (Street, City, State, Country)	Citizenship
Venkat Gopalan	Columbus, Ohio, USA	5220 Brandy Oaks Lane Columbus, Ohio 43210 USA	India
Signature:			Date:

Full Name (First, Middle, Last)	Residence Address (City, State, Country)	Post Office Address (Street, City, State, Country)	Citizenship
Milan Jovanovic	Dublin, Ohio, USA	6920 Sparrow Lane Dublin, Ohio 43235 USA	US
Signature:			Date:

Full Name (First, Middle, Last)	Residence Address (City, State, Country)	Post Office Address (Street, City, State, Country)	Citizenship
Paul S. Eder	Oreland, Pennsylvania, USA	312 Wishman Avenue Pennsylvania, 19075 USA	US
Signature:			Date:

Full Name (First, Middle, Last)	Residence Address (City, State, Country)	Post Office Address (Street, City, State, Country)	Citizenship
Tony Giordano	Phoenixville, Pennsylvania USA	38 North Spring Lane Phoenixville, Pennsylvania 19460 USA	US
Signature:			Date:

Full Name (First, Middle, Last)	Residence Address (City, State, Country)	Post Office Address (Street, City, State, Country)	Citizenship
Gordon D. Powers	Malvern,Pennsylvania USA	9 Lloyd Avenue Malvern, Pennsylvania 19355 USA	US
Signature:			Date:

Full Name (First, Middle, Last)	Residence Address (City, State, Country)	Post Office Address (Street, City, State, Country)	Citizenship
K. Asish Xavier	West Chester, Pennsylvania USA	1155 Kingsway Road #4 West Chester, Pennsylvania 19382 USA	India
Signature:			Date: